https://doi.org/10.1016/j.fuel.2023.127945

Machine Learning clustering algorithms for the automatic generation of Chemical Reactor Networks from CFD simulations

Matteo Savarese[a,b,1], Alberto Cuoci[c], Ward De Paepe[b], Alessandro Parente[a]

[a]*Aero-Thermo-Mechanics department, Université Libre de Bruxelles, Belgium*
[b]*Thermal Engineering and Combustion Unit, University of Mons (UMONS), Belgium*
[c]*CRECK Modeling Lab, Department of Chemistry, Materials, and Chemical Engineering, Politecnico di Milano*

Abstract

Predicting the thermal and environmental performances of combustion systems can be difficult and computationally expensive. Chemical Reactor Networks (CRN) represent an appealing solution for performing faster numerical simulations since they can carry out calculations with detailed kinetics in a significantly reduced amount of time. However, the design of such models is challenging, as it usually requires a considerable amount of expertise from the user. In this work, we present a novel, automatic methodology for the design of CRN models, which consists of the post-processing of CFD data via a combination of unsupervised clustering and graph scanning algorithms. The methodology was tested on a semi-industrial furnace which can operate in Moderate or Intense Low oxygen Dilution (MILD) combustion. The furnace operates at a nominal power input of 15 kW and is fed with several CH_4-H_2 mixtures, with two different air injector diameters. First, RANS simulations of the different cases were performed and validated against experimental data. Subsequently, different CRNs were extracted singularly for each case and simulated using detailed kinetics mechanisms. The different CRNs showed good performances in predicting NO emissions for the entire range of cases, as the results were in reasonable agreement with experimental data. Last, the capability of a single CRN to extrapolate towards different operating

[1] Corresponding author
Email address: matteo.savarese@ulb.be (Matteo Savarese)

conditions was also assessed by using one single CRN for the simulation of different fuel mixtures. The approach showed a good level of generalization since the NO predictions were close to the experimental values.

Keywords: MILD combustion, Chemical Reactor Networks, Machine Learning

---

1. Introduction

Climate change and global warming have become a serious concerns for our society. As a result, in the last few years, many policies have been implemented to limit anthropogenic greenhouse gases emission, pushing our economy to become completely carbon-neutral by 2050. Renewable Energies (RE) will play a significant role in the transition towards a net-zero greenhouse gas emission energy system, and their share in the global energy mix is expected to grow exponentially over the next decades [1]. To ensure a secure, stable, and economically viable energy supply, technological challenges to cope with the intermittent nature of RE must be addressed. Among possible alternatives, Power-To-Fuel represents one of the most promising solutions for a long-term and cost-effective RE storage strategy, since the eventual energy surplus can be stored in the form of synthetic fuels, that can be shipped and used on-demand [2]. In this context, hydrogen is emerging as a leading Smart Energy Carrier (SEC), since it can be directly employed as a fuel with zero emissions of CO_2, delivering high-quality energy in a secure and cost-effective manner [3]. However, if proper combustion technologies are not employed, hydrogen can lead to an increased production of harmful pollutants (i.e. NO_x), formed during high-temperature processes, along with other undesired phenomena, due to its increased reactivity with respect to traditional fossil fuels. For this reason, breakthroughs in combustion technologies are required to handle carbon-free fuels in a clean, safe, and efficient manner.

Numerical models and simulations can be used as strategic tools to assist in designing novel combustion assets, allowing for the screening and testing of new concepts while reducing trivial and expensive experimental campaigns. In this framework, the combustion modelling community massively relies on Computational Fluid Dynamics (CFD) for its ability to predict turbulent reactive flow fields even in complex industrial geometries. Despite major advances in this field, predicting low-concentration

pollutants emissions (e.g. NO_x and CO) remains challenging due to prohibitive computational requirements associated with CFD simulations involving a large number of chemical species [4]. For those reasons, alternative and more computationally-affordable modelling tools for emissions predictions have been developed.

Chemical Reactor Networks (CRN) undoubtedly represent one of the most appealing solutions to perform simulations with detailed kinetics with a consistent saving in computational requirements. This approach, first introduced by Bragg [5], allows one to schematically describe a complex flow field by dividing the combustor domain into a discrete number of compartments. The selected zones are then modelled as canonical chemical reactors, namely Perfectly Stirred Reactor (PSR) or Plug Flow Reactor (PFR), so that conservation equation for energy and species are solved only in a few blocks. The reactors are interconnected, exchanging mass and heat, thus forming a network that is intended to preserve the main features of the entire flow field. As a result, designing an equivalent CRN model of a given system is not an easy task, since the performances of the model are highly affected by the adopted network configuration. The network structure can be selected *a priori* based on flow considerations, as suggested by Andreini and Facchini [6] to perform parametric emissions studies on typical gas turbine combustors, or by Yousefian et al. [7, 8] to perform uncertainty quantification of operating conditions in lean premixed combustion. The authors reported a good agreement between model predictions and experimental data. On the other hand, the *a priori* design requires a highly experienced user and a considerable amount of prior knowledge about the system.

CFD can greatly help to retrieve important features of the flow field, and to incorporate them in the CRN structure. In several works [9, 10, 11, 12, 13] CRNs were designed by manually observing the results of CFD simulations, identifying the relevant regions in the combustor by looking at temperature and velocity fields. Other works introduced a more systematic zonal division by grouping the computational CFD cells based on temperature and local stoichiometry [14, 15], and, following this methodology, integrated CFD-CRN tools were developed by Falcitelli et al. [16, 17]. A sophisticated Kinetic Post Processor (KPP), able to automatically group CFD computational cells into a large number of reactors (up to $\sim 10^4$) was developed by Cuoci et al. [18, 19] and improved by Stagni et al. [20]. A similar tool was used by Monaghan et al. [21] on

a CH_4 diffusion flame. Fichet et al. [22] introduced the concept of fluid age in the clustering process. Recently, Perpignan et al. [23] developed a more systematic methodology for grouping the computational cells based on a graph transversal algorithm, namely Breadth First Search. However, in those works, a large number of reactors is employed ($\sim O(10^2)$), making those approaches computationally quite demanding. Employing a low number of reactors can result in a tool able to give instantaneous predictions, as shown by Kaluri et al. [24], who predicted lean-blowout in real-time, and a CRN-based control system to prevent it was developed on the same test case [25]. However, the CRN comprising 3 reactors was designed *a priori* based on previous knowledge of the system. An extensive review of the use of CRN in combustion modelling can be found here [26], but it is clear that advances are needed to make the design of CRN more efficient and automatic. Recently

Machine Learning (ML) can represent an attractive tool to help extract the flow field's key features from CFD simulations data. ML is receiving increasing attention from the combustion community for its ability to manipulate huge amounts of data and extract knowledge from them. Unsupervised clustering algorithms have been widely employed to discover patterns in combustion data by grouping observations exhibiting similar features. Clustering based on Local Principal Component Analysis (LPCA) [27] was employed by Parente et al. for the identification of low-dimensional manifolds in MILD combustion [28]. Coussement et al. [29] used LPCA to develop a more efficient tabulation approach for combustion simulations based on PC transport. LPCA was also used for dynamic chemistry reduction methods [30, 31] and for the interpretation of both DNS [32], and LES data of MILD combustion simulations [33]. The kmeans algorithm [34], which is probably the most popular routine for clustering tasks, was employed by Yu et al. [35] to identify various soot formation regions for Reactivity Controlled Compression Ignition (RCCI) engine mode. The same algorithm was used by Zhang et al. [36], combined with dimensionality reduction, to identify combustion regimes in turbulent non-premixed flames. Therefore, it is clear that using Machine Learning can help reduce the amount of time and user expertise required for designing cost-effective and reliable CRN models. For those reasons, we aim to extend the use of clustering algorithms for CRN applications.

In this work, an automatic and integrated methodology for efficient CRN design based on unsupervised learning has been developed. In particular, clustering approaches were employed to post-process CFD data of a MILD-capable semi-industrial furnace for the automatic partition of the computational domain into macro zones that exhibit similar thermo-chemical features to generate CRN models. In addition to the unsupervised clustering step, the CRN extraction algorithm was further extended, including graphs objects to guarantee that the final output of the clustering process represents geometrically connected zones in the computational domain. CRN simulations of the different networks were conducted with detailed kinetics using the software NetSMOKE++ [37]. Moreover, the capabilities of a single CRN model to extrapolate towards different operating conditions were also assessed.

The paper is organized as follows: Section 2 provides details regarding the system under study and describes the CFD approach used for the simulations. Section 3 describes the different steps of the post-processing methodology, while in Section 4, the results obtained with both CFD and CRN simulations are reported.

2. Test Case: ULB furnace

2.1. Experimental System

The selected test case is a semi-industrial, MILD-capable furnace available at the Universit´e Libre de Bruxelles. The combustion chamber, schematically represented in Figure 1, is cubic shaped, with 1000 mm length for each side, including a 300 mm thick insulation layer made of ceramic fibre, to limit heat losses through the surrounding. The fuel and the oxidizer are injected at the bottom of the furnace, and the hot flue gases flow out co-axially at the same location. A recuperative heat exchanger is mounted at the bottom, which allows preheating of the air so that the reactive mixture is above its self-ignition temperature after the initial mixing between fuel and oxidizer. The internal fluid-dynamics design of the furnace provides the recirculation of exhaust products within the reaction region, which is fundamental to achieving the MILD combustion regime. Inside the chamber, four finned air-cooled tubes are present to simulate an external thermal load, which can be adjusted by tuning the air flow rate.

The system can be fed with CH_4 - H_2 mixtures in variable proportions at a nominal thermal input of up to 20 kW. Customization of the air injector diameter is also possible, thus allowing for more flexibility in the internal fluid pattern.

Different fueling mixtures ranging from 0% H_2 to 100% H_2 by volume were investigated. The selected equivalence ratio and the thermal power input were 0.8 and 15 kW, respectively. The combustion of the mixtures was analyzed with two different air injector diameters with an internal diameter of 16 and 25 mm. The experimental cases examined are summarized in Table 1.

Experimental measurements available from previous works [38, 39] include temperature measurements sampled in different locations of the chamber and chemical composition sampled at the outlet, including minor species such as NO_x. The heat dissipation values, $\dot{Q}_{loss}$, were calculated considering the heat losses through radiation, conduction through the wall and the cooling system. A more detailed explanation can be found here [39].

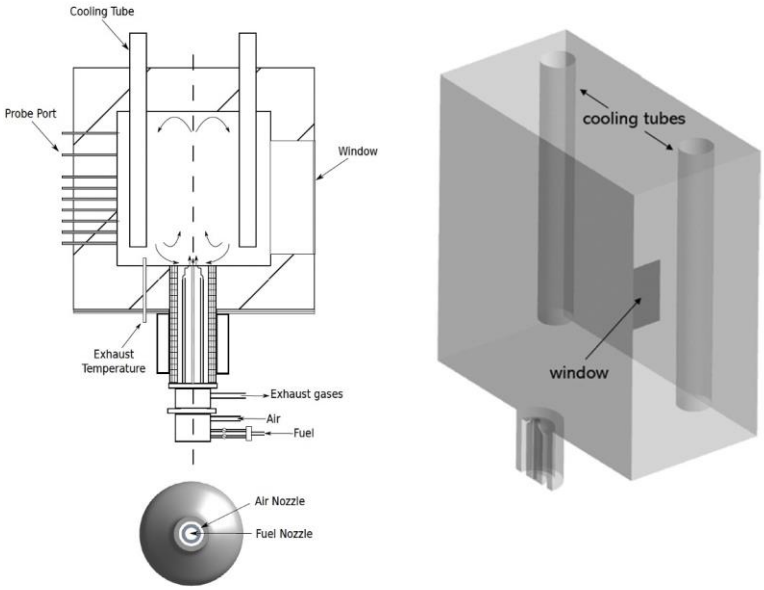


Figure 1: Schematic representation of the furnace on a 2-D cross-section (left) and on a 3-D view (right)

2.2 CFD modeling

2.2.1 Numerical modeling approach

Reynolds-Averaged Navier-Stokes (RANS) based simulations of the system were carried out by employing the commercial software Ansys Fluent 19.3®. The standard $k-\varepsilon$ model was used for turbulence modelling. Since MILD conditions are encountered in this work, finite-rate chemistry effects were accounted for using the Partially-Stirred Reactor (PaSR) model [40]. This model relies on the hypothesis that the computational cell is split into a reactive and a non-reactive part. The reactive fraction of the cell, κ, can be expressed as:

$$\kappa = \frac{\tau_c}{\tau_c + \tau_{mix}} \tag{1}$$

Cases examined (16 mm and 25 mm ID)				
CH_4/H_2 [vol%]	$\dot{m}_{air}$ [g/s]	$\dot{m}_{fuel}$ [g/s]	T_{fuel} [K]	$\dot{Q}_{loss}$ [kW]
100/0	6.4	0.3	340	11.0
75/25	6.32	0.284	340	11.1
50/50	6.2	0.259	340	11.3
25/75	5.9	0.217	340	11.0
0/100	5.3	0.125	340	11.0

Table 1: Operating conditions considered for the CFD simulations, each of the cases performed with an air injector diameter of 16 mm and 25 mm

where τ_c and τ_{mix} represent the chemical and mixing timescales, respectively. The mean, chemical reactive source term of the generic species *s* can then be expressed as:

$$\overline{\dot{\omega}}_s = \kappa \dot{\omega}_s^*(\widetilde{Y}, \widetilde{T}) \tag{2}$$

Where $\dot{\omega}_s^*(\tilde{Y}, \tilde{T})$ is the formation rate of the species *s* in the reactive structure, which is modelled as an ideal reactor, and the associated ODE system is solved using the selected chemical mechanism over a fixed residence time [41]. $\tilde{Y}$ and $\tilde{T}$ are the Favre-

Averaged mass fractions and temperature, respectively. Several methods to estimate the chemical and mixing timescale are available in the literature [40, 42]. In the present work, the chemical timescale was evaluated directly from the species formation rate [40], while the mixing timescale was calculated using a constant C_{mix} [42] value of 0.5:

$$\tau_{mix} = C_{mix}\frac{k}{\varepsilon}. \tag{3}$$

The Kee chemical mechanism (17 species, 34 reactions) [43] without NO formation was used to model chemistry to retain enough chemistry details and keep the computational cost on an affordable level at the same time. Radiation effects were accounted for using the Discrete Ordinate model with the Weighted-Sum-Of-Grey-Gases for the calculation of the absorption properties of the mixture, using the coefficients proposed by Smith et al. [44].

To include also information regarding the residence time distribution inside the combustor, an additional transport equation for local residence time was solved, following the approach of Fichet et al. [22]. The scalar quantity corresponding to the non-reactive, non-diffusive local residence time of the fluid, τ_f, can be seen as a scalar transported by convection which evolves with a linear dependence on time, such that the derived transport equation can be written as [45]:

$$\frac{D\tau_f}{Dt} = 1 \tag{4}$$

In this way, information regarding the local residence time can be subsequently exported and included in the clustering process.

Additionally, to get information about the temperature fluctuations inside the furnace, which can strongly influence NO formation [46], an estimation of the temperature oscillations magnitude is required. Therefore, an additional transport equation for the temperature variance, $\widetilde{T''^2}$, can be solved as follows:

$$\frac{\partial}{\partial t}\left(\rho\widetilde{T''^2}\right) + \nabla\cdot\left(\rho\mathbf{v}\widetilde{T''^2}\right) = \nabla\cdot\left(\frac{\mu_t}{\sigma_t}\nabla\widetilde{T''^2}\right) + C_g\mu_t(\nabla\widetilde{T})^2 - C_d\rho\frac{\varepsilon}{k}\widetilde{T''^2} \tag{5}$$

where ρ is the density, $\mathbf{v}$ is the velocity vector, μ_t is the turbulent viscosity while σ_t,C_g and C_d are model constants. In equation 5, σ_t represents the turbulent Prandtl number, which assumes a value of 0.85 [47]. The second term on the right-hand side of the equations is the production of scalar flux due to mean gradients, where the constant C_g

is equal to 2.86 [46, 47]. The last term on the right-hand side represents the scalar dissipation rate, and C_d here assumes a value of 2.0 [47, 48].

2.2.2 Computational domain and boundary conditions

The simulations were carried out on a simplified, 2-D, structured grid to limit the CPU requirements. In particular, two different axisymmetric grids were used for the two different air injectors' diameters, containing roughly 45k and 35k computational cells for the air injectors of 16 and 25 mm, respectively. The grid selection was based on a mesh sensitivity study. Regarding the fuel and air inlets, a mass flow inlet was provided. The temperature of the fuel is directly available from experiments, while the temperature of the preheated air was estimated by an energy balance [39]. The thermal losses through the environment were modelled by prescribing a fixed negative heat flux at the walls. Its value was calculated by considering the heat losses detected experimentally due to convection and radiation [39]. The values of mass flow rates, inlet temperatures and overall heat losses can be found in Table 1.

3. Post-processing methodology

The post-processing methodology involves several steps, which can be described as follows:

1. Data Preparation
2. Application of an unsupervised clustering algorithm, namely *k*-means
3. Re-assignment of the data points by means of a graph scanning algorithm to enforce the spatial connectivity of the clusters
4. Adjustment of the mass flow rates between the clusters to guarantee perfect closure of the mass balance
5. CRN simulation of the generated network with the correction of kinetic constants to take into account temperature fluctuations

A detailed description of each step is reported below.

3.1. Data preparation

Once a CFD simulation is complete, the data stored in the computational cells are exported. More specifically, the obtained dataset contains the thermochemical state of each computational cell (e.g. temperature and mass fractions) and can be indicated as **X**. The data matrix **X** has than *m* rows and *p* columns, where *m* is the number of computational cells and *p* the number of selected variables, thus each row of the matrix can be expressed as a vector $\mathbf{x_i} = [T^i, Y^i_{CH_4}, Y^i_{O_2}, \dots]$.

At this point, a standard procedure to center and scale the data to be able to compare quantities in different units of measure is carried out. The data are commonly centered by subtracting the mean of each variable, while several options are available for scaling. In general, a centered and scaled point can be expressed as:

$$\tilde{x}_{i,j} = \frac{x_{i,j} - \bar{x}_j}{d_j} \tag{6}$$

where $\bar{x}_j$ and d_j represent the mean and the scaling factor for the $j{-}th$ variable, respectively. In this work, the data were scaled using the *range* criterion, where $d_j = \max(\mathbf{x}_j) - \min(\mathbf{x}_j)$.

Regarding the set of exported variables to perform clustering, a subset of features from the CFD solution is selected. It is important to include features that represent both the thermo-chemical state of the system and its fluid dynamics regime. Temperature and a reaction progress variable are selected to be representative of the thermo-chemical composition space. The progress variable is defined as the sum of H_2O, CO_2 and CO mass fractions over their equilibrium value, as:

$$f = \frac{Y_{\mathrm{H_2O}} + Y_{\mathrm{CO_2}} + Y_{\mathrm{CO}}}{Y^{eq}_{\mathrm{H_2O}} + Y^{eq}_{\mathrm{CO_2}} + Y^{eq}_{\mathrm{CO}}} \tag{7}$$

The selected fluid dynamic variables are velocity magnitude, velocity angle, and local residence time (expressed as in equation 4). In conclusion, the final data matrix is composed of five features and around 30k observations for the 25 mm case and 45k for the 16 mm injector case.

3.2. Unsupervised Clustering

Clustering is a popular task in data analysis, which consists in grouping unlabelled objects that show similar features, or characteristics, in order to find patterns in the data. There exists a large variety of algorithms, which are applied in many scientific areas [49]. Many available algorithms aim to partition the dataset's points into a prescribed number of clusters so that a given cost function is minimized. The *k*−means algorithm, which belongs to the latter category, is probably one of the most common and simplest algorithms used in exploratory analysis. The idea is to assign the points of the dataset into a user-defined number of clusters *k*, such that the sum of squared distances between the points and the clusters' centroids is minimized. In other words, the objective function can be defined as:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{m} ||\mathbf{x_i^{(j)}} - \mathbf{c_j}||^2 \tag{8}$$

where $||\mathbf{x_i^{(j)}} - \mathbf{c_j}||^2$ is the euclidean distance between the *i−th* point belonging to the *j−th* cluster and the cluster centroid $\mathbf{c_j}$. The optimization problem is solved through a so-called Loyd type routine [34], as follows:

1. **Initialization**: *k* random centroids are initialized
2. **Partitioning**: the data points are assigned to the cluster with the nearest centroid
3. **Update**: the new *k* centroids are calculated by performing the arithmetic mean of the observations in each cluster
4. **Covergence**: If convergence criteria are met, the algorithm stops, otherwise steps 2 and 3 are repeated iteratively

One drawback of this routine is that the algorithm can stagnate around a local minimum during the solution of the optimization problem. Therefore, the algorithm is quite sensitive to its initialization. On the other hand, in this work reproducibility of results is important, therefore uniform initialization with a fixed random generator was selected. In particular, the MATLAB® implementation of *k*−means was used in this work.

The choice of a suitable number of clusters is a contentious topic in the framework of unsupervised learning, as it is usually selected using heuristic or trialand-error

procedures. As an alternative, statistical metrics can provide useful *a priori* insights for appropriate values of *k*. Many statistical metrics are available in the literature, with the most popular being the silhouette index [50] or the Davies-Bouldin index [51]. In practice, those metrics asses how well the clusters are separated (extra-cluster distance) and how similar the points in a certain cluster are (intra-cluster distance or cluster's dispersion), in terms of Euclidean distance. Among solutions with different number of clusters, the value of *k* that shows largest separation and smallest dispersion should be picked. The silhouette index calculate an affinity measure between 0 and 1 for each data point, which makes the calculation of such index quite tedious if the dataset is large. On the other hand, the Davies-Bouldin index calculates the overall average ratio between intra-cluster distance and extra cluster distance, as follows:

$$DB = \frac{1}{k}\sum_{i=1}^{k}\max_{j\neq i}\left(\frac{\bar{d}_i + \bar{d}_j}{d_{i,j}}\right) \tag{9}$$

where *k* is the number of clusters, $\bar{d}_i$, $\bar{d}_j$ are the averages distance within cluster *i* and *j*, respectively, while $d_{i,j}$ is the distance between the centroids of cluster *i* and *j*. As a result, the most suitable value of *k* is the one that minimizes the value of the index. In this work, the minimization procedure was used to select an appropriate number of clusters for the post-processing approach. However, since here clustering is used to derive a physics-based model, namely CRNs, an *a posteriori* study is also required, to assess the effect of *k* from a model predictions perspective, and it was carried out in this work.

3.3. Geometrical connectivity extension

For CRN applications, spatially homogeneous areas within the domain need to be identified. Unsupervised clustering algorithms do not include any information about the spatial connectivity of the data points, so the obtained clustering solutions do not guarantee that the identified clusters are spatially connected. Including Cartesian coordinates (*x*,*y*,*z*) in the clustering dataset can promote spatial connectivity without providing formal assurance that the resulting clusters are geometrically connected. To include information about points connectivity, graph objects can be employed. A graph *G* is a set of nodes connected by edges, such that $G = G(V,E)$, where *V* are the nodes and *E* are the edges. The whole computational mesh can be represented as a graph, where the computational cells represent the nodes and their connection represent the

edges. The mesh graph is originally entirely connected, implying that starting from a certain point V_i, there exists a path along the edges by which it is possible to reach any other node V_j. Once an unsupervised learning algorithm is applied to the data, the cell centres (namely the nodes of the graphs) are labelled with their corresponding cluster index. A schematic infographic is shown in Figure 2. The points highlighted in different colours represent the identified clusters via unsupervised learning. As we can observe, those clusters do not represent entirely connected graphs, but they are divided in multiple unconnected components. To guarantee the spatial connectivity of the zones, the clusters' subgraphs are scanned, and the unconnected pieces are re-assigned. More specifically, unconnected graphs larger than 1% of the total volume of the physical domain are split, and new clusters (or new areas) are added. Smaller clusters are then re-assigned to their immediate neighbour, more specifically, the closest bordering cluster in terms of Euclidean distance, such that physical connectivity is enforced. Moreover, to avoid possible numerical instabilities to the CRN solver, clusters that occupy a physical volume below 0.01% of the global volume are eliminated, and their points are assigned to their immediate neighbour in terms of Euclidean distance. This is done since highly reduced volumes result in very small values of residence time in the reactor, which require a fine pseudo-transient time step for the ODE solver. In addition, the presence of a few small reactors result in more connections between the reactors. Due to the sequential nature of the solver, having more connections results in more iterations. As a consequence, the convergence of the global system can be considerably slowed down.

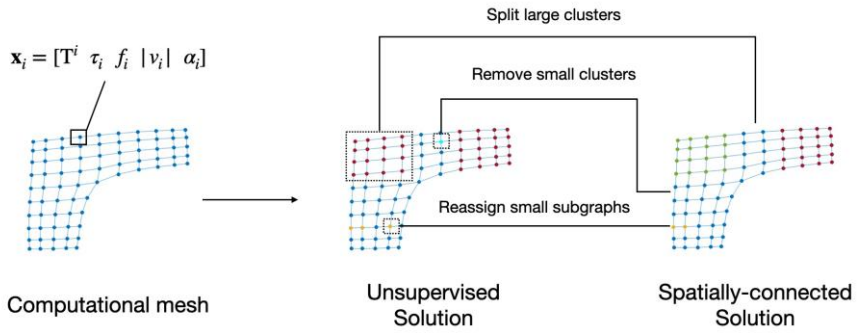


Figure 2: Graphs and connectivity scanning process

3.4. Controlling the mass imbalance

Due to rounding errors, CFD computational cells are often characterized by a small amount of mass imbalance. As a result, a certain amount of such imbalance can also be present across the identified clusters. The numerical solution of the ODEs associated with the CRN model requires the mass flow rates to be perfectly closed to satisfy the conservation of mass. Therefore, the mass imbalance across different clusters must be considered and kept under a certain threshold. In this work, the relative mass imbalance is evaluated for each reactor. A correction procedure is applied if the imbalance is above a selected threshold for at least one of the reactors. The relative mass imbalance for the *i*-th reactor is calculated as the difference between the input and output mass flow rates, normalized by the maximum value among the two. The correction procedure is applied if more than 1% of imbalance is detected in at least one of the reactors, and it can be described as follows.

First, the mass flow rates exchanged between the reactors obtained at the end of the clustering are stored in a squared matrix **M**, with dimension$N_R \times N_R$, where N_R is the number of reactors and the element M_{ij} is the mass flowing from reactor *i* to reactor *j*. Second, the total outlet mass flowrate of the *i*-th reactor can be calculated as $M_{out}^{i} = \sum_{j=1}^{N_R} M_{ij}$. Then, the fraction of the outlet mass from reactor *i* to reactor *j* is calculated as $\alpha = M_{ij}/M_{out}^{i}$ and a matrix $\boldsymbol{\alpha}$ is obtained. The mass balance for each reactor can be rewritten in a matrix form as

$$\mathbf{M_{out}} - \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{M_{out}} = \mathbf{b_{ext}}, \tag{10}$$

where $\mathbf{M_{out}}$ and $\mathbf{b_{out}}$ are the $N_R \times 1$ vector corresponding to the outlet mass flows from flows from the reactors and external source of outputs and inputs, respectively Equation 10 can be rewritten as:

$$(\mathbf{I} - \boldsymbol{\alpha}^{\mathrm{T}})\mathbf{M_{out}} = \mathbf{b_{ext}}, \tag{11}$$

where **I** is the $N_R \times N_R$ identity matrix. This equation corresponds to a linear system in the form $\mathbf{Ax} = \mathbf{b}$, where finding **x** corresponds to solving for updated $\mathbf{M_{out}}$ that actually close the mass balance of the network.

3.5. CRN modeling

Each identified compartment can be treated as a canonical chemical reactor model, occupying a volume equal to the volume of the corresponding cluster. Since clustering algorithms identify zones with similar thermo-chemical features, the zones are modelled as Perfectly Stirred Reactors (PSR). A PSR is a 0-D chemical reactor model, which implies the hypothesis of perfectly mixed reactants inside the control volume so that the flow is not resolved [52]. A network of PSR consists of interconnected reactors, which exchange mass and energy. The problem statement is finding the thermo-chemical status, namely temperature and chemical composition in each reactor, given the boundary conditions (inlet conditions and thermal exchange), the volumes of the reactors and the mass exchanged between them. The mass exchanged can be stored in a matrix **M**, where each element M_{ij} represents the convective mass flux from reactor i to reactor j. For each reactor, we can write N_S conservation equations, where N_S is the number of species:

$$\rho_i V_i \frac{dY_k^i}{dt} = \sum_{j=1}^{N_R} M_{ji}\left(Y_k^j - Y_k^i\right) + \dot{\Omega}_k^i V_i + \sum_{b=1}^{N_{bc}} M_b\left(Y_k^b - Y_k^i\right) \quad \forall i = 1, ..., N_R \quad (12)$$

where ρ_i, V_i and Y_k^i are the gas density, the volume and the mass fraction of the $k{-}th$ species in the $i{-}th$ reactor, respectively. M_b and Y_k^b are the mass flow and the mass fractions of external sources of mass, given by the boundary conditions, while $\dot{\Omega}_k^i$ is the net production rate of the $k{-}th$ species. Additionally, an equation for the conservation of energy can be solved:

$$\rho_i \hat{c}_{P,i} V_i \frac{dT_i}{dt} = \sum_{j=1}^{N_R} M_{ji}\left(\hat{h}_j - \hat{h}_i\right) - \sum_{k=1}^{N_S} h_k^i \dot{\Omega}_k^i V_i + \sum_{b=1}^{N_{bc}} M_b\left(\hat{h}_b - \hat{h}_i\right) + \dot{Q}_{ext} \quad \forall i = 1, ..., N_R \quad (13)$$

where T_i is the temperature of the $i{-}th$ reactor and $\widehat{c_p}$ the specific heat of the mixture. $\hat{h}$ represents the mass-weighted enthalpy per unit of mass in a reactor and $\dot{Q}_{ext}$ is the heat exchanged by the reactor with the environment (e.g. conduction, radiation...). The solution of the ODE system associated with the CRN is solved through a sequential solver, namely the NetSMOKE++ [37] solver, which is an extension of the software OpenSMOKE++ developed by Cuoci et al. [53]. The software solves the conservation equations sequentially for each reactor (one at a time). It calculates the final residuals

of the global mass balance for the chemical elements (e.g. C, N, O...). At each iteration, the values of Y_k^i and T_i are updated until final convergence is reached.

With this approach, it is possible to employ a chemical mechanism that differs from the one used to perform the CFD simulations. In particular, more detailed kinetic schemes can be employed to include all possible formation pathways of minor pollutants.

3.6. Effect of temperature oscillations and average reaction rate evaluation

During the CRN post-processing calculations, the average temperature in each reactor needs to be evaluated. It can be directly calculated from the solution of the energy balance (equation 13) if non-isothermal reactors are considered or can be estimated from CFD by calculating the average temperature in each cluster. In both methods, especially in the isothermal case, the calculated temperature is an estimation of the average temperature. However, in the case of turbulent combustion, calculating reaction rates only as a function of the average temperature does not consider the possible effect of temperature fluctuations on the kinetic constants. Due to the strong non-linear coupling between temperature and reaction rates, the impact of those fluctuations on the net formation rates of the chemical species can be high. The average kinetic rate of an Arrhenius-like reaction, $k(T) = A_0 T^{\beta} \exp\left(-\frac{E_a}{RT}\right)$, differs from the kinetic rate calculated at the average temperature:

$$\overline{k}(T) \neq k(\overline{T}) \tag{14}$$

Following the approach of Cuoci et al. [46], the averaged kinetic rate can be calculated by integrating the kinetic rate over a presumed probability density function (PDF) of the temperature distribution, as:

$$\overline{k}(T) = \int_{T_{min}}^{T_{max}} k(T)p(T)dT \tag{15}$$

where T_{min} and T_{max} are the extremes of the PDF and $p(T)$ is the value of the PDF.

The integral in Equation 15 is carried out by considering the normalized temperature, $\xi = \frac{T - T_{min}}{T_{max} - T_{min}}$, where T_{min} and T_{max} were chosen in such a way that a sufficiently large domain of integration is allowed. In this work, T_{min} was selected by removing 100 K from the minimum temperature encountered in the CFD simulation, and T_{max} as the

maximum temperature plus 500 K, since typically higher fluctuations are detected in high temperature regions.

Therefore, Equation 15 can be rewritten as:

$$\overline{k}(\xi) = \int_0^1 k(\xi)p(\xi)d\xi \tag{16}$$

In this work, a β-PDF distribution for the temperature is assumed, as this type of PDF is considered more accurate in the context of combustion [48]. The mean and the variance of the PDF are estimated from CFD, considering the average temperature $\tilde{T}$ and the temperature variance $\widetilde{T''^2}$ described before in equation 5. During the calculation, the averaged value of the kinetic rate is evaluated for each reaction at each iteration, thus accounting for the effect of temperature oscillations.

4. Results

4.1. CFD simulations

RANS simulations of the experimental cases were carried out. In particular, each fuel composition was simulated with the two different air injectors, and temperature predictions were compared with experimental data sampled at different locations within the furnace. Results obtained with the air injector of 16 mm are reported in Figure 3. Predictions are in good agreement with experimental data, especially at axial locations z > 200 mm. For the case with pure H_2, the temperature is slightly overpredicted close to the burner (z = 100 mm). This can be explained by the fact that when H_2 is present at high concentrations (> 75% by volume), the flame is attached at the burner exit, which makes it difficult to obtain reliable experimental measurements. The temperature is also slightly overpredicted in the case with 25% H_2 by volume. This is because this is a transitional case between the conventional and the MILD regime, which is thus more challenging to model. The regime transition is caused by an increased H_2 concentration in the fuel. More specifically, since H_2 is more reactive than CH_4, it reacts at a location closer to the burner exit, where the entrainment of flue gases is lower. When the H_2 content exceeds a certain threshold, oxygen dilution is no longer

high enough to sustain MILD regime, and more pronounced temperature gradients are observed [38].

Results for the injector diameter of 25 mm are reported in Figure 4. A good agreement

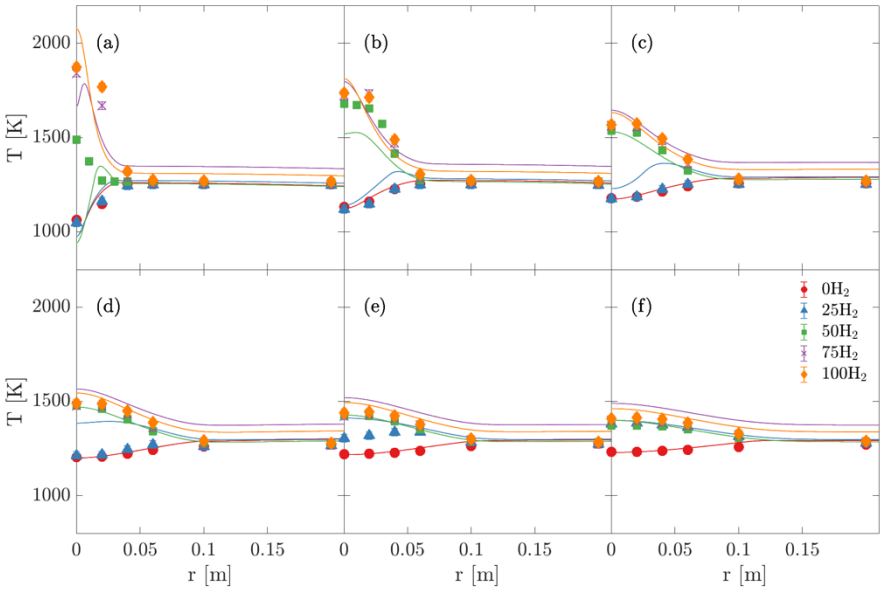


Figure 3: Comparison of temperature predictions (lines) with experimental data (points) along the radial directions for some selected axial locations, z = 100 mm (a), z = 200 mm (b), z = 300 mm (c), z = 400 mm (d), z = 500 mm (e) and z = 600 mm (f) with an air injector diameter of 16 mm

is found between experiments and predictions. As for the 16 mm case, H_2 temperature is slightly overpredicted but within less than 10% of relative error with respect to the experimental value. For this injector, the predictions for the transitional case are improved. Overall, the CFD model was able to capture correctly the temperature profile detected experimentally within a 10% relative error bound with respect to the experimental measurements.

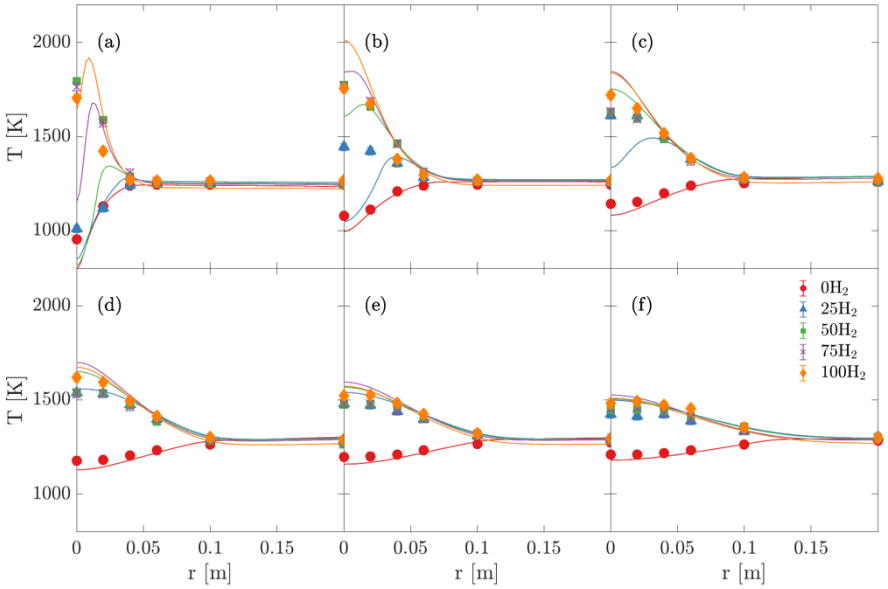


Figure 4: Comparison of temperature predictions (lines) with experimental data (points) along the radial directions for some selected axial locations, z = 100 mm(a), z = 200 mm (b), z = 300 mm (c), z = 400 mm (d), z = 500 mm (e) and z = 600 mm (f) with an air injector diameter of 25 mm

4.2. Post processing results

4.2.1. Qualitative analysis of the clustering

A typical clustering output is reported in Figure 5 to qualitatively assess the unsupervised clustering algorithm. It can be observed that the clustering can differentiate the fuel and the oxidizer inlets. Moreover, a few small clusters are present within the combustor's central zone, where the flame is anchored. In that area, the highest variability of quantities of interest, e.g. temperature and species, is found, so several clusters are needed to capture that variability. One large cluster is found outside of the flame, representing a thermally homogeneous, post-flame area. The reassignment through graph analysis of the computational cells might worsen the unsupervised clustering solution since the latter is supposed to represent the optimal solution from a cost function minimization perspective. To investigate this aspect more quantitatively, the

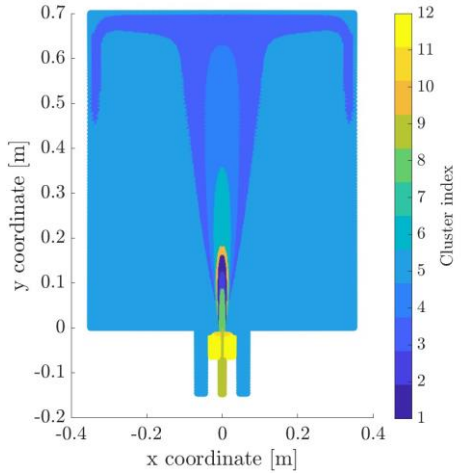


Figure 5: Typical output of the clustering procedure (*k*-Means plus graph analysis) on a 50-50% CH_4/H_2 case with an injector diameter of 25 mm

Davies-Bouldin (DB) index [51] was used. The DB index represents the ratio between the intra-cluster and extra-cluster distance for a given clustering solution. Therefore, the one with a lower DB should be picked when comparing several solutions. For the same case of Figure 5, namely the 50-50% CH_4-H_2 with air injector of 16 mm, *k*−Means was applied for several initial numbers of clusters *k*. The DB index was calculated before and after the graph reassignment, and results are reported in Figure 6. We can observe that there is no substantial difference between the DB value obtained before and after the graph reassignment. Moreover, the reassignment also positively affects the DB value, especially for a high number of clusters ($k > 20$). This can be explained by the fact that when the number of clusters is increased, many small clusters that are not particularly distinct from each other are created in the unsupervised step. Those clusters may include cells in different locations, far from each other, resulting in poor spatial homogeneity. The graph reassignment helps in re-joining neighbour cells that were separated by the unsupervised clustering.

It is important to consider that the final number of clusters can be different from the initial value of *k* selected for the unsupervised step because of

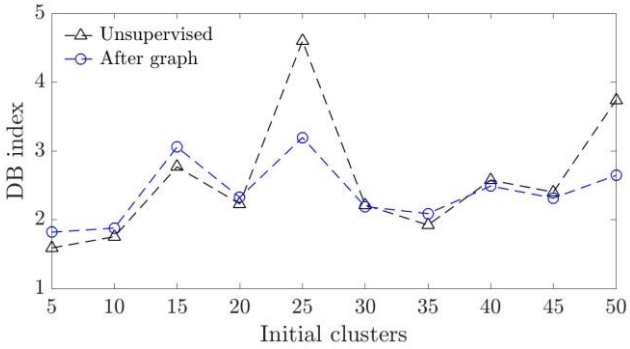


Figure 6: Comparison of DB index obtained before and after the graph reassignment for the 50-50 CH_4-H_2 case with the 16 mm injector.

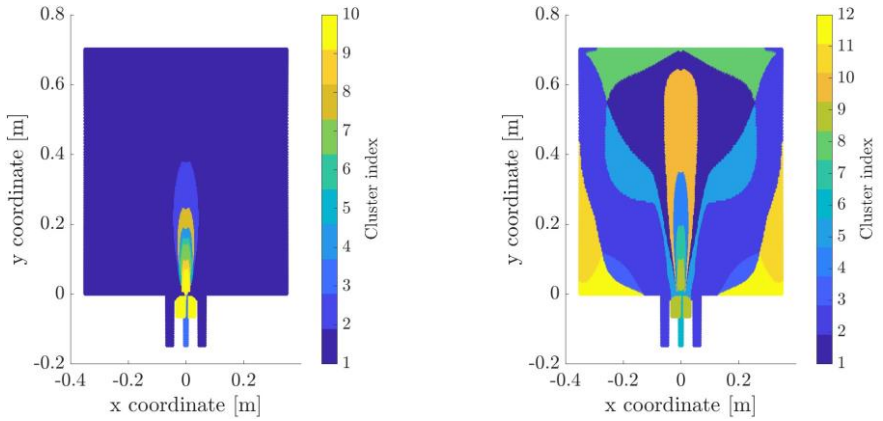
the graph reassignment. In this analysis (Figure 6), no volume threshold removal was applied since the scope of this step was to assess the capability of the graph re-assignment process of gaining a spatially-connected solution from the unsupervised step. In the next section, the effect of the number of clusters is analyzed from a model performance perspective, and the volume removal threshold is applied.

4.2.2. Qualitative analysis of different variables selections

In this work, a reduced set of variables was selected as reference set. The choice was made in such a way that the variable set included information of the fluid dynamic configuration of the system, the time history of the fluid particles and the most important thermo-chemical features. However, the effect of using different set of variables was investigated from a qualitative point of view, as it can highlight interesting physical aspects of the problem.

To test the sensitivity of the clustering, two different sets of variables were analyzed. The first set includes temperature and all the chemical species, without any information of the flow. The second set, in contrast, contains only the velocity magnitude and the velocity angle. For brevity, they will be referred as Set1 and Set2, respectively, while the reference set described in section 3.2 will be referred as Set0.

The clustering output for Set1 and Set2 using a starting value of k = 10 is reported in Figure 7. As we can observe, Set1 output (Figure 7a) is quite coherent with the results



(a) (b)

Figure 7: Comparison between clustering solutions obtained with two different sets of variables. (a) Temperature and chemical species mole fractions (Set1) and (b) velocity magnitude and velocity angle (Set2)

from Set0 (Figure 5), since many small clusters are present within the flame region. However, the evolution of the jet is not totally captured, resulting in a post-flame cluster which is probably excessively large. On the other hand, Set2 shows some deeply different results. As we can see from Figure 7b, the clusters follow the evolution of the expanding jet, and they highlight areas with different velocities orientation and location. From a CRN perspective, Set2 does not represent an optimal solution, as wider clusters are present within the flame region. As a consequence, the variability of temperature and chemical species in the flame clusters can be large, and the hypothesis of Perfectly Stirred Reactor can be compromised.

Summarizing, a good balance between flow variables and thermo-chemical variables, namely Set0, showed the best results in terms of a qualitative analysis.

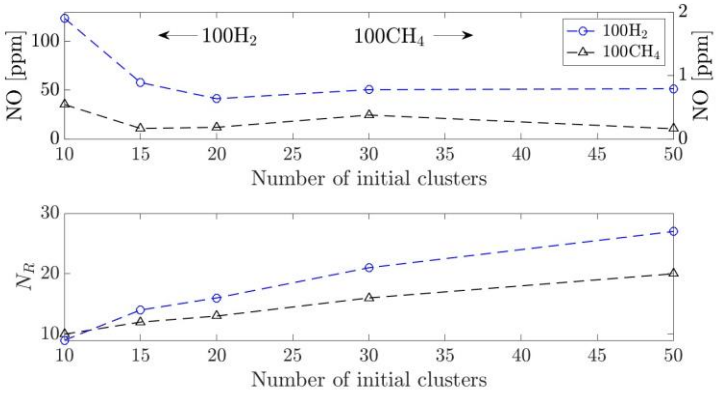


Figure 8: Effect of the number of reactors NO emissions (up) and the effective number of reactors used in the simulations (down) as a function of the initial clusters selected in the unsupervised step

4.3. CRN results

4.3.1. Effect of the number of reactors

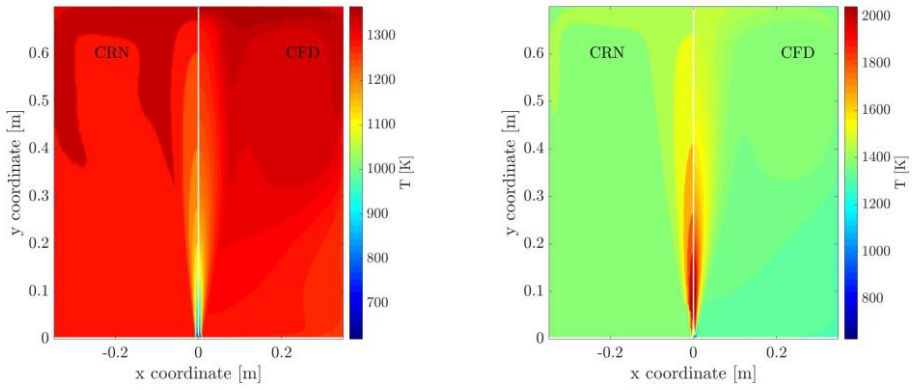
To assess the effect of the number of reactors from a model output perspective, CRN simulations with different numbers of reactors were performed on two selected test cases, namely the pure CH_4 and the pure H_2 cases on the 16 mm ID. The cases will be referred to as the $100CH_4$ case and the $100H_2$ case.

CRN simulations were performed using the POLIMI C1-C3 chemical mechanism [54] with different numbers of reactors, obtained by progressively increasing the number of initial clusters. It should be stressed for clarity that the number of reactors differs from the number of initial clusters selected, due to the volume threshold removal. In particular, as mentioned before, clusters smaller than the 0.01% of the global volume of the system are removed since they are not representative of the macro features of the system, and the points in those small clusters are re-assigned to their closest neighbour in terms of Euclidean distance. The number of initial clusters, for both the $100CH_4$ and the $100H_2$ cases, was varied from 10 to 50, and results of the analysis are reported in Figure 8. As we can observe, the solution for the $100CH_4$ case is more or less constant across the whole range of numbers of reactors, being less than 1 absolute ppm. For the $100H_2$ case, instead, the solution stabilizes where 20 initial clusters (16 reactors) are selected. Those cases were selected strategically for the differences in the combustion regime between the two. In fact, combustion in the $100CH_4$ case happens in pure MILD

conditions, with very smooth evolution of temperature and low absolute peaks, resulting in the absence of NO. This regime is known to approach a PSR from a reactive point of view. Therefore, a reduced number of reactors is required for its accurate chemical description. On the other hand, the $100H_2$ case happens in the conventional flame regime due to the enhanced reactivity of the mixture. Therefore, steeper temperature gradients and higher absolute temperature peaks are observed, requiring more ”discretization” in terms of reactors for these conditions to be captured correctly. This is further testified by the fact that when the number of initial clusters is increased, a higher number of reactors is obtained for the $100H_2$ case with respect to the $100CH_4$ case. Summarizing, the combustion regime influences the number of reactors needed from a CRN perspective. As a general rule, less reactors are required for MILD conditions, while more reactors are required for conventional flames, due to steeper gradients and higher temperature variability.

Contours of temperature for the 100 CH_4 and the 100 H_2 cases obtained with 13 and 17 reactors, respectively, which correspond to a starting number of clusters equal to 20, are also reported in Figure 9. As we can observe, the structure of the clustering can correctly represent the flow field’s main features, as it naturally follows the fluid dynamic structure of the system. Contours of a major species, namely H_2O, are also reported in Figure 10. It can be observed that the CRN model, despite its simple structure, is not only able to correctly predict the outlet value of H_2O, which comes from the stoichiometry of the main combustion reaction, but it is also able to follow its evolution in the system. This is important since it is directly related to a correct estimation of the residence time of the reactors.

It should be pointed out that the major scope of this work is to develop reducedorder, grey-box models for the simulation of realistic combustion assets. Therefore, the developed CRNs in each case should be representative of the macro-



(a) (b)

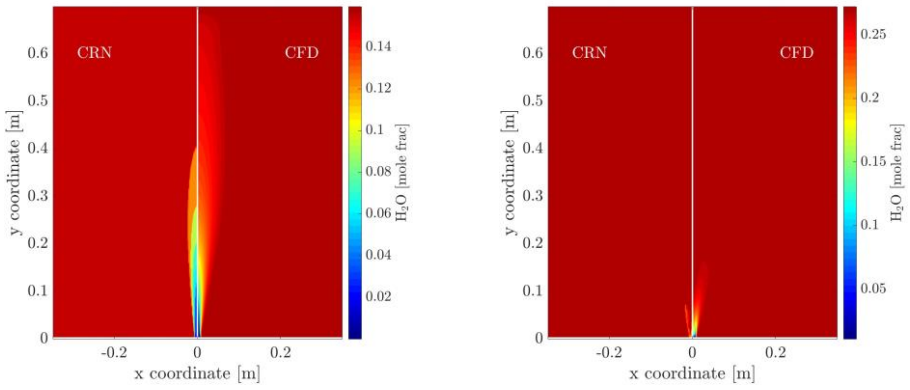
Figure 9: Comparison between temperature contours of CFD and CRN for the (a) CH_4 case and (b) the H_2 case on the 16 mm ID obtained with 20 initial clusters, corresponding to 13 reactors for the $100CH_4$ case and 17 reactors for the $100H_2$ case

features of the system and not a high-fidelity replica of the CFD simulations.

Details of the computational requirements of the CRN simulations as a function of the number of reactors are reported in Figure 11. The simulations were carried out on a local machine, and the chemical mechanism employed, namely the Polimi C1-C3, included 149 species and 2459 reactions, and it was the largest employed in this work, therefore characterized by the highest CPU requirements. As we can observe, the time required for the CRN simulations to converge is a linear function of the number of reactors, and takes only a few minutes, even seconds for highly schematic CRNs (e.g. $N_R = 10$). In contrast, each CFD simulation of this particular test case can take up to 24 hours of running time, in parallel on 20 cores. Therefore, the computational saving is considerable.

4.3.2. Results for the whole fuel composition range

The CRN methodology was then applied to a range of CFD simulations (Table 1) by setting 20 as the initial number of clusters for each case, considering the results from the previous section. Different kinetic mechanisms were used in this step to assess the influence of the chemical mechanism on the network



(a) (b)

595 Figure 10: Comparison between H_2O contours of CFD and CRN for the (a) CH_4 case and (b) the H_2 case on the 16 mm ID obtained with 20 initial clusters, corresponding to 13 reactors for the $100CH_4$ case and 17 reactors for the $100H_2$ case

predictions. In particular, the GRI 2.11 [55], and the San Diego [56] kinetic mechanisms were used. Results on the 16 mm and the injector are reported in Figure
600 12, where normalized NO emissions are compared with experimental measurements. A good overall agreement between experimental data and CRN predictions can be observed. In general, NO predictions for pure CH_4 and very little amount of H_2 in the fuel are in excellent agreement with experimental data. Regarding kinetic mechanisms, the GRI 2.11 tends to slightly overestimate the NO predictions when the H_2 content
605 increases. The POLIMI and the San Diego mechanisms offer better performances than the GRI 2.11, considering their more complex structure. Moreover, the CRN post-processor developed in this work outperformed the default post-processor available in Ansys Fluent R19.3®.

The CRN post-processor appears versatile towards different regimes, fuel mixtures and
610 jet configurations, providing a reasonable estimation of NO emissions across the whole case range. This testifies to the ability of unsupervised learning to automatically identify a reduced number of important flow field features, namely the identified compartments, without requiring a high level of expertise from the user.

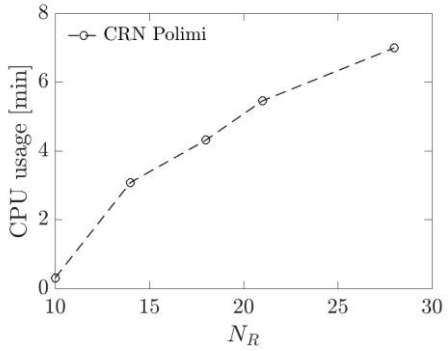
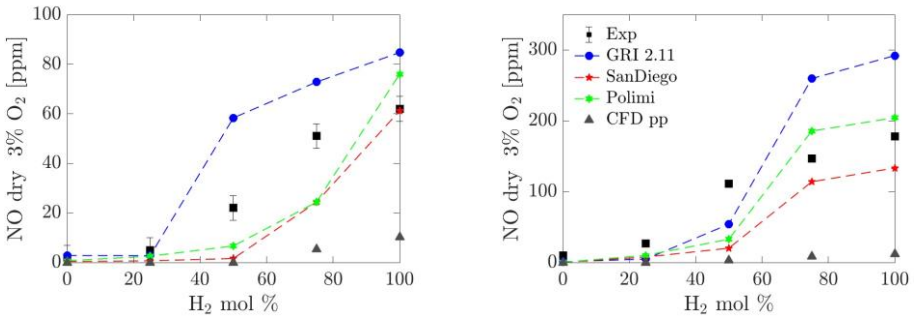


Figure 11: Computational cost of CRN simulations using the NetSMOKE++ solver as a function of the number of effective reactors.

4.4. Generalizability of the model

The post-processing technique showed good results when a network of isothermal reactors was extracted in every single case. However, the possibility of constructing a more general model that can be used for a wider range of operating conditions appears more appealing and also quite challenging. In particular, via unsupervised learning, it was possible to identify a reactor network structure for each case that could predict NO in good agreement with experimental data, but to what extent can this structure also correctly predict emissions from other cases? Therefore, the capabilities of this approach to generalize towards a wider range of operating conditions were tested. In particular, for both injectors, the intermediate 50-50% CH_4-H_2 case was selected, and a CRN was extracted. Differently from the previous sections, non-isothermal reactors must be used in this analysis to be able to use a CRN with different boundary conditions, as temperature from CFD may not always be available.

To further reduce the user supervision, the number of reactors was selected by performing the optimization of the Davies-Bouldin (DB) index (eq. 9), on the 16 mm ID and the 25 mm ID. To perform this task, the number of clusters was varied parametrically, and the solution showing the minimum DB index



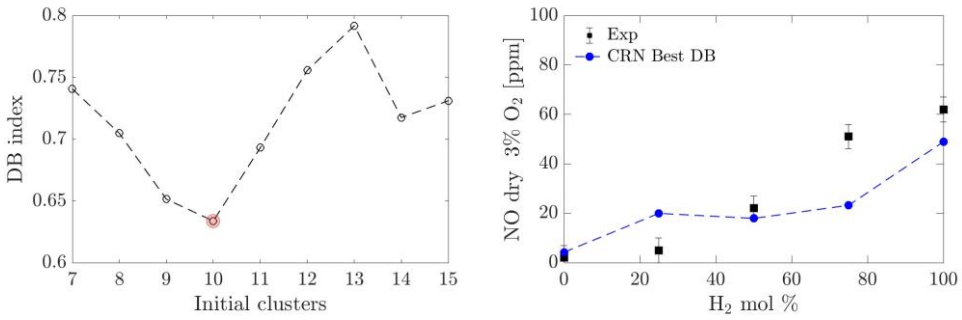
(a) (b)

Figure 12: Comparison of NO emissions between experiments and CRN predictions on the (a) 16 mm injector and the (b) 25 mm injector diameters. CRN were extracted on each case by selecting 20 initial clusters. "CFD pp" stands for the CFD post-processor of Ansys Fluent R19.®

was selected. To calculate the internal mass flow rates, since the boundary conditions change by changing the fuel mixture, the splitting ratios were calculated for the CRN extracted on the 50-50% CH_4-H_2 case and equation 10 was solved for $\mathbf{M_{out}}$ with the new boundary conditions vector $\mathbf{b_{ext}}$. Regarding the temperature fluctuations amplitude estimation, the ratio between temperature oscillations magnitude and the average temperature was calculated in each cluster on the 50-50% CH_4-H_2 case and applied to the other cases. Finally, a global heat transfer coefficient was prescribed for the largest post-flame reactor to model heat losses to match the average outlet temperature, which is common across all the cases examined.

Results of this analysis for the 16 mm case are shown in Figure 13. According to the DB analysis, the optimal CRN structure was found for k = 10, which corresponded to 10 reactors. As we can observe from Figure 13b, the identified CRN could predict NO emissions for all the fuel mixtures with a reasonable agreement with experimental data. The San Diego chemical mechanism was used in this analysis, as it showed the best performances in the previous sections. The 25-75% H_2-CH_4 is slightly overpredicted, which can be attributed to the discrepancy between CFD and the experimental temperature field, with the latter being slightly overpredicted. The 75-25% and the 100-0% H_2-CH_4 cases appear underpredicted by the CRN, which can be due do the rough estimation of the temperature oscillations. Results for the 25 mm are also reported, in

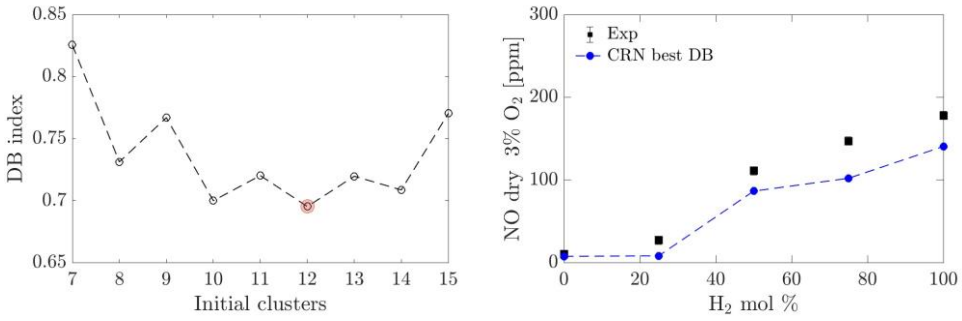
Figure 14. Also, in this case, predicted NO values are in good agreement with experimental data.



(a) (b)

Figure 13: (a) Parametric analysis of the DB index with the number of clusters and (b) NO predictions on the whole fuel mixtures range using the solution which minimizes the DB index for the 16 mm injector case

In this case, as shown in Figure 14a, the optimal CRN structure was found for 12 initial clusters, which corresponded exactly to 12 reactors. However, the difference between the DB value from $k = 10$ and $k = 12$ is pretty much negligible, thus meaning that those cases are almost identical. This result is in line with the DB analysis carried out on the 16 mm injector. This means that the two injectors' configurations show similarities between their macro-structure of the flow field. On the other hand, the different jet Reynolds number influences the amount of flue gases recirculated in the reaction region. Therefore, it is necessary to use different networks for the two injectors, which are similar in terms of clustering shape but show differences in the internal mass flow rates.



(a) (b)

Figure 14: (a) Parametric analysis of the DB index with the number of clusters and (b) NO predictions on the whole fuel mixtures range using the solution which minimizes the DB index for the 25 mm injector case

5. Conclusions and perspectives

A novel methodology based on unsupervised clustering and graph analysis capable of extracting equivalent Chemical Reactor Networks models of realistic combustion devices from CFD data was developed. The method was applied on a *quasi*-industrial test case, namely a MILD-capable furnace fed with CH_4/H_2 mixtures with two different air injectors diameters. CFD simulations were carried out using reduced kinetics to limit the computational efforts, while detailed kinetic mechanisms were used to simulate the extracted CRN models. RANS simulations showed good performances for several cases, thus confirming the capability of the PaSR model to model the combustion of different fuels at different operating conditions and regimes. Regarding the CRN post-processing procedure, some main conclusions can be drawn:

- The unsupervised clustering algorithm was able to automatically detect important regions within the combustor domain, identifying inlets, inflame reactors and post-flame reactors
- The graph scanning algorithm further extended the capabilities of unsupervised algorithms without affecting the *goodness* of the identified unsupervised solution
- CRN results showed good predictions and promising results for all the cases examined in this work, providing close-to-experiment NO emissions

- The CRN extracted on the intermediate cases for the two injectors showed the possibility for the model to generalize towards different fuel compositions. This means that the main features of the flow field can be captured by unsupervised learning algorithms, reducing the supervision required for constructing well-performing CRN

- Overall, the approach is promising as it has the potential to drive towards the development of reduced-order, physics-based combustion models without a huge amount of expert-user supervision

Although some aspects of the methodology can be improved, this work showed the potential of unsupervised learning towards synthesising reliable grey-box models, which can be of key importance for combustion monitoring and control. Moreover, this approach has the potential to become a general framework for the automatic generation of compartment-based models for many other chemicalengineering-related problems.

Acknowledgement

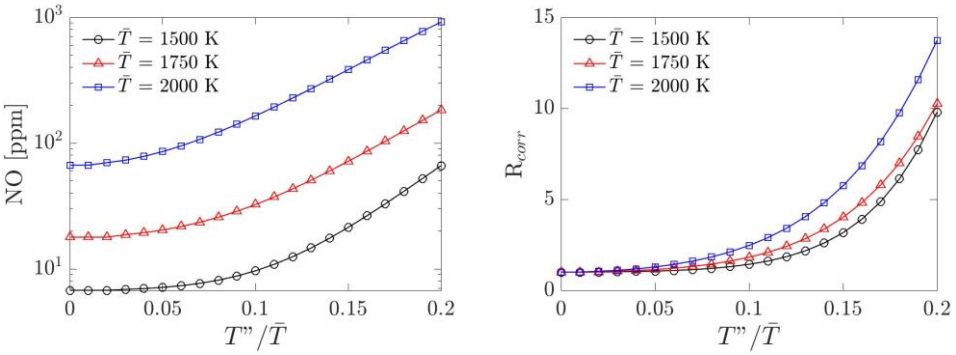
The present received funding from the Fonds National de la Recherche Scientifique (FNRS) of Belgium and by the Energy Transition Fund of Belgium. The project was carried out under the framework of the European Union's Horizon 2020 research and innovation programme under grant agreement No. 714605 (funded by the European Research Council).

We want to thank professor Tiziano Faravelli from Politecnico di Milano for the fruitful discussions and precious advice and Himanshu Dave from Universit´e Libre de Bruxelles for helping in the shape of the work.

Appendix A: Effect of kinetic corrections on NO formation

Temperature oscillations can strongly impact the formation of some selected species, especially those characterized by rather slow chemistry. As explained in the work of Cuoci et al. [46], the chemical reactions that are mainly influenced by temperature oscillations are characterized by high activation energy. As a result, NO_x, especially the ones which are formed via the thermal pathway [57], are more sensitive to this aspect. Therefore, the method introduced in Section 3.6 can have a strong influence on the final CRN predictions regarding NO emissions.

In this section, the effect of average temperature and oscillation amplitude is discussed. CRN simulations are carried out on a highly simplified configuration, namely an isothermal PSR, to isolate the effects of temperature oscillations. The reactor is fed with pure H_2 and air at an equivalence ratio of 0.8, and the residence time (τ_R) is set to 0.01 s, which is similar to the actual residence time within the flame region estimated for the furnace under study. The effect of temperature oscillation parameters, namely average temperature and temperature fluctuations amplitude (T''), is shown in Figure 15.



(a) (b)

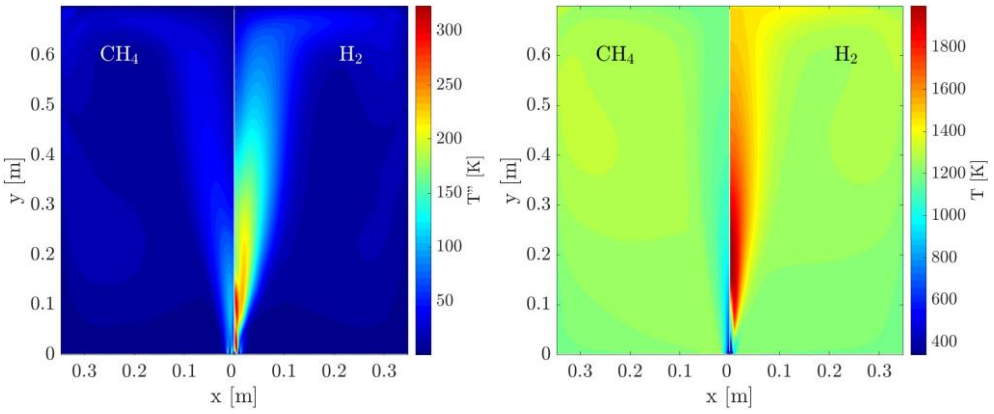
Figure 15: (a) Overall NO emissions from a PSR (τ_R = 0.01 s) and (b) ratio between NO predicted including the kinetic correction and NO predicted without kinetic corrections (R_{corr}) as a function of temperature oscillations amplitude, for three different average temperatures.
The fuel is pure H_2 at equivalence ratio of 0.8

As we can observe from Figure 15a, there is a strong non-linear relationship between the normalized oscillations' amplitude ($T''/\bar{T}$) and overall NO formation, as the final

value of NO increases exponentially with the increase in temperature variance. The correction ratio R_{corr}, which is equal to the ratio between NO obtained with the kinetic correction and without, is shown in Figure 15b. Overall, the R_{corr} values obtained for the three different average temperatures are very similar, and with an increase of more than 10 times the uncorrected value when the temperature variance is 20% of the average temperature.

Regarding the furnace, the effect of kinetic corrections on the global NO emissions is expected to be more pronounced when high values of temperature variance are detected. Contours of the temperature variance and the absolute temperature for two deeply different fuel mixtures are reported in Figure 16. The results are extracted from the CFD simulation of pure CH_4 and pure H_2 for the 25 mm injector case. As shown in Figure 16a, significantly higher oscillations are detected for the pure H_2 case. Moreover, observing Figure 16b, the temperature is significantly higher for the H_2 case, with peaks above 1800 K, which is known to be a sort of threshold for the formation of thermal NO. Therefore, the cases at rather low temperatures, which are happening in MILD conditions, show considerably lower combustion temperatures and quasi-absent temperature oscillations, which then do not significantly impact NO formation. On the other hand, for conventional flame regime cases, not only is higher temperature detected, but the magnitude of temperature oscillations is also significantly higher, and this can have a significant impact on NO formation. In conclusion, temperature oscillations can amplify the final NO formation by a factor of up to 15 when high average temperature and high-temperature variance are detected. In particular, with increasing H_2 in the fuel, the combustion regime is shifted towards conventional flame operations, thus increasing both average temperature and temperature oscillations, with a considerable increase in overall NO formation.

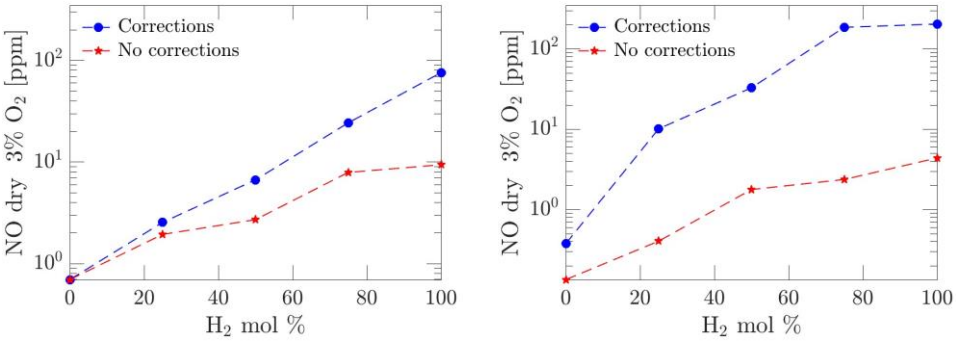
The comparison between NO predicted using the kinetic correction approach and without the corrections is reported in Figure 17. As we can observe, the influence of



(a) (b)

Figure 16: (a) Temperature fluctuation amplitude and (b) absolute temperature contours comparison between pure CH_4 and H_2 on the 25 mm injector case. Results taken from the CFD simulations

temperature fluctuations dramatically impacts the overall NO formation, especially for increased H_2 content in the fuel, when the combustion regime shifts towards the conventional flame.



(a) (b)

Figure 17: Comparison between NO emissions obtained with kinetic corrections (blue) and without kinetic corrections for the different fuel mixtures with the (a) 16 mm and the (b) 25 mm air injectors' diameter

References

[1] I. E. Agency, World energy outlook (2021).
URL www.iea.org/weo

[2] C. Wulf, P. Zapp, A. Schreiber, Review of power-to-x demonstration projects in europe, Frontiers in Energy Research 8 (2020) 191. doi: 10.3389/fenrg.2020.00191.

[3] F. Dawood, M. Anda, G. Shafiullah, Hydrogen production for energy: An overview, International Journal of Hydrogen Energy 45 (7) (2020) 3847– 3869. doi:https://doi.org/10.1016/j.ijhydene.2019.12.059.

[4] T. Lu, C. Law, Towards accomodating realistic fuel chemistry in largescale computations, Progress in Energy and Combustion Science 35 (2009) 192–215. doi:10.1016/j.pecs.2008.10.002.

[5] S. Bragg, Application of reaction rate theory to combustion chamber analysis, aeronautical Research Council London (1953).

[6] A. Andreini, B. Facchini, Gas Turbines Design and Off-Design Performance Analysis With Emissions Evaluation , Journal of Engineering for Gas Turbines and Power 126 (1) (2004) 83–91. doi:10.1115/1.1619427.

[7] S. Yousefian, G. Bourque, R. F. D. Monaghan, Uncertainty Quantification of NOx and CO Emissions in a Swirl-Stabilized Burner, Journal of Engineering for Gas Turbines and Power 141 (10). doi:10.1115/1.4044204.

[8] Y. Sajjad, G. Bourque, R. Monaghan, Bayesian inference and uncertainty quantification for hydrogen-enriched and lean-premixed combustion systems, International Journal of Hydrogen Energy 46 (46) (2021) 23927– 23942. doi:10.1016/j.ijhydene.2021.04.153.

[9] I. Novosselov, P. Malte, S. Yuan, R. Srinivasan, J. C. Y. Lee, Chemical reactor network application to emissions prediction for industial dle gas turbine, Proceedings of the ASME Turbo Expo 1. doi:10.1115/GT2006-90282.

[10] M. Kanniche, Coupling cfd with chemical reactor network for advanced

NO_x prediction in gas turbine, Clean Techn Environ Policy 12 (2010) 661– 670. doi:10.1007/s10098-010-0293-5.

[11] D. Lee, J. Park, J. Jin, M. Lee, A simulation for prediction of nitrogen oxide emissions in lean premixed combustor, Journal of Mechanical Science and Technology 25 (2011) 1871–1878. doi:10.1007/s12206-011-0425-9.

[12] A. A. V. Perpignan, M. G. Talboom, Y. Levy, A. G. Rao, Emission modeling of an interturbine burner based on flameless combustion, Energy & Fuels 32 (1) (2018) 822–838, pMID: 29910533. arXiv:https://doi.org/10.1021/acs.energyfuels.7b02473, doi:10. 1021/acs.energyfuels.7b02473.
URL https://doi.org/10.1021/acs.energyfuels.7b02473

[13] T. H. Nguyen, J. Park, C. Sin, S. Jung, S. Kim, Numerical investigation of the pressure effect on the nox formation in a lean-premixed gas turbine combustor, Energy & Fuels 35 (8) (2021) 6776–6784. doi:10.1021/acs. energyfuels.0c02909.

[14] D. Benedetto, S. Pasini, M. Falcitelli, C. L. Marca, L. Tognotti, NO_x emission prediction from 3-d complete modelling to reactor network analysis, Combustion Science and Technology 153 (1) (2000) 279–294. doi: 10.1080/00102200008947265.

[15] T. Faravelli, L. Bua, A. Frassoldati, A. Antifora, L. Tognotti, E. Ranzi, A new procedure for predicting nox emissions from furnaces, Computers & Chemical Engineering 25 (4) (2001) 613–618. doi:https://doi.org/10. 1016/S0098-1354(01)00641-X.

[16] M. Falcitelli, L. Tognotti, S. Pasini, An algorithm for extracting chemical reactor network models from cfd simulation of industrial combustion systems, Combustion Science and Technology 174 (11-12) (2002) 27–42. doi:10.1080/713712951.

[17] M. Falcitelli, S. Pasini, N. Rossi, L. Tognotti, Cfd + reactor network analysis: an integrated methodology for the modeling and optimisation of industrial systems

for energy saving and pollution reduction, Applied Thermal Engineering 22 (8) (2002) 971–979. doi:https://doi.org/10.1016/ S1359-4311(02)00014-5.

[18] A. Cuoci, A. Frassoldati, G. Buzzi Ferraris, T. Faravelli, E. Ranzi, The ignition, combustion and flame structure of carbon monoxide/hydrogen mixtures. note 2: Fluid dynamics and kinetic aspects of syngas combustion, International Journal of Hydrogen Energy 32 (15) (2007) 3486–3500, international Symposium on Solar-Hydrogen-Fuel Cells 2005. doi:https://doi.org/10.1016/j.ijhydene.2007.02.026.
URL https://www.sciencedirect.com/science/article/pii/S0360319907001255

[19] A. Cuoci, A. Frassoldati, A. Stagni, T. Faravelli, E. Ranzi, G. BuzziFerraris, Numerical modeling of nox formation in turbulent flames using a kinetic post-processing technique, Energy & Fuels 27 (2) (2013) 1104– 1122. doi:10.1021/ef3016987.

[20] A. Stagni, A. Cuoci, A. Frassoldati, T. Faravelli, E. Ranzi, A fully coupled, parallel approach for the post-processing of cfd data through reactor network analysis, Computers & Chemical Engineering 60 (2014) 197–212. doi:10.1016/j.compchemeng.2013.09.002.

[21] R. F. D. Monaghan, R. Tahir, A. Cuoci, G. Bourque, M. Fu¨ri, R. L. Gordon, T. Faravelli, A. Frassoldati, H. J. Curran, Detailed multi-dimensional study of pollutant formation in a methane diffusion flame, Energy & Fuels 26 (3) (2012) 1598–1611. doi:10.1021/ef201853k.

[22] V. Fichet, M. Kanniche, P. Plion, O. Gicquel, A reactor network model for predicting nox emissions in gas turbines, Fuel 89 (9) (2010) 2202–2210. doi:https://doi.org/10.1016/j.fuel.2010.02.010.
URL https://www.sciencedirect.com/science/article/pii/S0016236110000591

[23] A. A. V. Perpignan, R. Sampat, A. Gangoli Rao, Modeling pollutant emissions of flameless combustion with a joint cfd and chemical reactor network approach, Frontiers in Mechanical Engineering 5. doi:10.3389/fmech. 2019.00063.
URL https://www.frontiersin.org/articles/10.3389/fmech.2019. 00063

[24] A. Kaluri, P. Malte, I. Novosselov, Real-time prediction of lean blowout using chemical reactor network, Fuel 234 (2018) 797–808. doi:10.1016/ j.fuel.2018.07.065.

[25] S. Gupta, P. Malte, S. Brunton, I. Novosselov, Prevention of lean flame blowout using a predictive chemical reactor network control, Fuel accepted. doi:10.1016/j.fuel.2018.09.044.

[26] H. Khodayari, F. Ommi, Z. Saboohi, A review on the applications of the chemical reactor network approach on the prediction of pollutant emissions, Aircraft Engineering and Aerospace Technology ahead-of-print. doi:10. 1108/AEAT-08-2019-0178.

[27] N. Kambhatla, T. Leen, Dimension reduction by local principal component analysis, Neural Computation 9 (1997) 1493–1516. doi:10.1162/neco. 1997.9.7.1493.

[28] A. Parente, J. Sutherland, L. Tognotti, P. Smith, Identification of low-dimensional manifolds in turbulent flames, Proceedings of the Combustion Institute 32 (1) (2009) 1579–1586. doi:https://doi.org/10.1016/j.proci.2008.06.177.
URL https://www.sciencedirect.com/science/article/pii/ S154074890800299X

[29] A. Coussement, O. Gicquel, A. Parente, Mg-local-pca method for reduced order combustion modeling, Proceedings of the Combustion Institute 34 (1) (2013) 1117–1123. doi:https://doi.org/10.1016/j.proci. 2012.05.073.
URL https://www.sciencedirect.com/science/article/pii/ S1540748912000740

[30] G. D'Alessio, A. Parente, A. Stagni, A. Cuoci, Adaptive chemistry via pre-partitioning of composition space and mechanism reduction, Combustion and Flame 211 (2020) 68–82. doi:https://doi.org/10.1016/j.combustflame.2019.09.010.
URL https://www.sciencedirect.com/science/article/pii/S0010218019304249

[31] G. D'Alessio, A. Cuoci, G. Aversano, M. Bracconi, A. Stagni, A. Parente, Impact of the partitioning method on multidimensional adaptive-chemistry simulations, Energies 13 (10). doi:10.3390/en13102567.
URL https://www.mdpi.com/1996-1073/13/10/2567

[32] H. Dave, N. Swaminathan, A. Parente, Interpretation and characterization of mild combustion data using unsupervised clustering informed by physics-based, domain expertise, Combustion and Flame 240 (2022) 111954. doi:https://doi.org/10.1016/j.combustflame.2021.111954.
URL https://www.sciencedirect.com/science/article/pii/S0010218021006970

[33] Z. Li, S. Tomasch, Z. X. Chen, A. Parente, I. S. Ertesv˚ag, N. Swaminathan, Study of mild combustion using les and advanced analysis tools, Proceedings of the Combustion Institute 38 (4) (2021) 5423–5432. doi:https://doi.org/10.1016/j.proci.2020.06.298.
URL https://www.sciencedirect.com/science/article/pii/S1540748920303904

[34] S. Lloyd, Least squares quantization in pcm, IEEE Transactions on Information Theory 28 (2) (1982) 129–137. doi:10.1109/TIT.1982.1056489.

[35] W. Yu, F. Zhao, W. Yang, H. Xu, Integrated analysis of cfd simulation data with k-means clustering algorithm for soot formation under varied combustion conditions, Applied Thermal Engineering 153 (2019) 299–305. doi:https://doi.org/10.1016/j.applthermaleng.2019.03.011.
URL https://www.sciencedirect.com/science/article/pii/S1359431118349172

[36] H. Zhang, H. Lu, F. Xie, T. Ma, X. Qian, Combustion regime identification in turbulent non-premixed flames with principal component analysis, clustering and back-propagation neural network, Processes 10 (8) (2022) 1653.

[37] S. Trespi, H. Nicolai, P. Debiagi, J. Janicka, A. Dreizler, C. Hasse, T. Faravelli, Development and application of an efficient chemical reactor network model for oxy-fuel combustion, Energy & Fuels 35 (9) (2021) 7121–7132. arXiv:https://doi.org/10.1021/acs.energyfuels.0c03560, doi:10.1021/acs.energyfuels.0c03560.
URL https://doi.org/10.1021/acs.energyfuels.0c03560

[38] M. Ferrarotti, W. De Paepe, A. Parente, Reactive structures and nox emissions of methane/hydrogen mixtures in flameless combustion, International Journal of Hydrogen Energy 46 (68) (2021) 34018–34045. doi:https://doi.org/10.1016/j.ijhydene.2021.07.161.
URL https://www.sciencedirect.com/science/article/pii/S0360319921028597

[39] M. Ferrarotti, M. Fu¨rst, E. Cresci, W. de Paepe, A. Parente, Key modeling aspects in the simulation of a quasi-industrial 20 kw moderate or intense low-oxygen dilution combustion chamber, Energy & Fuels 32 (10) (2018) 10228–10241. doi:10.1021/acs.energyfuels.8b01064.
URL https://doi.org/10.1021/acs.energyfuels.8b01064

[40] Z. Li, M. Ferrarotti, A. Cuoci, A. Parente, Finite-rate chemistry modelling of non-conventional combustion regimes using a partially-stirred reactor closure: Combustion model formulation and implementation details, Applied Energy 225 (2018) 637–655. doi:https://doi.org/10.1016/j.apenergy.2018.04.085.
URL https://www.sciencedirect.com/science/article/pii/S0306261918306433

[41] Z. Li, A. Cuoci, A. Sadiki, A. Parente, Comprehensive numerical study of the adelaide jet in hot-coflow burner by means of rans and detailed chemistry, Energy 139 (2017) 555–570.
doi:https://doi.org/10.1016/j.energy.2017.07.132.

URL https://www.sciencedirect.com/science/article/pii/S0360544217313142

[42] M. Ferrarotti, Z. Li, A. Parente, On the role of mixing models in the simulation of mild combustion using finite-rate chemistry combustion models, Proceedings of the Combustion Institute 37 (4) (2019) 4531–4538. doi:https://doi.org/10.1016/j.proci.2018.07.043.
URL https://www.sciencedirect.com/science/article/pii/S1540748918304619

[43] R. Bilger, S. St˚arner, R. Kee, On reduced mechanisms for methane-air combustion in nonpremixed flames, Combustion and Flame 80 (2) (1990) 135–149. doi:https://doi.org/10.1016/0010-2180(90)90122-8.
URL https://www.sciencedirect.com/science/article/pii/0010218090901228

[44] T. F. Smith, Z. F. Shen, J. N. Friedman, Evaluation of Coefficients for the Weighted Sum of Gray Gases Model, Journal of Heat Transfer 104 (4) (1982) 602–608. arXiv:https://asmedigitalcollection.asme.org/ heattransfer/article-pdf/104/4/602/5563512/602_1.pdf, doi:10. 1115/1.3245174.

URL https://doi.org/10.1115/1.3245174

[45] F. Ghirelli, B. Leckner, Transport equation for the local residence time of a fluid, Chemical Engineering Science 59 (3) (2004) 513–523. doi:https://doi.org/10.1016/j.ces.2003.10.013.
URL https://www.sciencedirect.com/science/article/pii/S0009250903005311

[46] A. Cuoci, A. Frassoldati, G. Buzzi Ferraris, T. Faravelli, E. Ranzi, The ignition, combustion and flame structure of carbon monoxide/hydrogen mixtures. note 2: Fluid dynamics and kinetic aspects of syngas combustion, International Journal of Hydrogen Energy 32 (15) (2007) 3486–3500, international Symposium on Solar-Hydrogen-Fuel Cells 2005. doi:https://doi.org/10.1016/j.ijhydene.2007.02.026.

URL https://www.sciencedirect.com/science/article/pii/S0360319907001255

[47] ANSYS, Inc., ANSYS Fluent Theory Guide, Canonsburg, PA 15137 (2013).

[48] R. O. Fox, Computational Models for Turbulent Reacting Flows, Cambridge Series in Chemical Engineering, Cambridge University Press, 2003. doi:10.1017/CBO9780511610103.

[49] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, C.-T. Lin, A review of clustering techniques and developments, Neurocomputing 267 (2017) 664–681. doi:https://doi.org/10.1016/j.neucom.2017.06.053.
URL https://www.sciencedirect.com/science/article/pii/S0925231217311815

[50] A. Dudek, Silhouette index as clustering evaluation tool, in: Conference of the Section on Classification and Data Analysis of the Polish Statistical Association, Springer, 2019, pp. 19–33.

[51] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (2) (1979) 224–227. doi:10.1109/TPAMI.1979.4766909.

[52] O. Levenspiel, Chemical Reaction Engineering, John Wiley and Sons, New York, 1999, 3rd edition.

[53] A. Cuoci, A. Frassoldati, T. Faravelli, E. Ranzi, Opensmoke++: An object-oriented framework for the numerical modeling of reactive systems with detailed kinetic mechanisms, Computer Physics Communications 192 (2015) 237–264. doi:https://doi.org/10.1016/j.cpc.2015.02.014.
URL https://www.sciencedirect.com/science/article/pii/S0010465515000715

[54] E. Ranzi, C. Cavallotti, A. Cuoci, A. Frassoldati, M. Pelucchi, T. Faravelli, New reaction classes in the kinetic modeling of low temperature oxidation of n-alkanes,

Combustion and Flame 162 (5) (2015) 1679–1691. doi:https://doi.org/10.1016/j.combustflame.2014.11.030.
URL https://www.sciencedirect.com/science/article/pii/S0010218014003848

[55] C. Bowman, R. Hanson, D. Davidson, W. G. Jr., V. Lissianski, G. Smith, D. Golden, M. Frenklach, M. Goldenberg, Gri 2.11 chemical mechanism.
URL http://combustion.berkeley.edu/gri-mech/new21/version21/text21.html#cite

[56] Chemical kinetic mechanism for combustion applications, Mechanical and Aerospace Engineering (Combustion Research) University of San Diego California (USA), chemical mechanism web page.
URL https://web.eng.ucsd.edu/mae/groups/combustion/mechanism. html

[57] Y. B. Zeldovich, The oxidation of nitrogen in combustion and explosions, Physicochimica U.S.S.R. 21 (1946) 577–628.